Prueba hipótesis

Manuel Francisco Romero Ospina

2023-03-17

Iniciamos cargando los paquetes y librerías para trabajar prueba hipótesis:

```
#install.packages("BSDA")
#install.packages("tigerstats")
library(tigerstats)
library("BSDA")
library(readr)
```

Prueba Hipótesis para una y dos muestras

Carga de base de datos

Vamos a trabajar la base denominada StudentsPerformance.csv.

```
df <- read_csv("StudentsPerformance.csv")
head(df,4)
```

```
## # A tibble: 4 x 8
##   gender `race/ethnicity` parental level~1 lunch test ~2 math ~3 readi~4 writi~5
##   <chr>  <chr>            <chr>            <chr> <chr>     <dbl>   <dbl>   <dbl>
## 1 female group B          bachelor's degr~ stan~ none         72      72      74
## 2 female group C          some college     stan~ comple~      69      90      88
## 3 female group B          master's degree  stan~ none         90      95      93
## 4 male   group A          associate's deg~ free~ none         47      57      44
## # ... with abbreviated variable names 1: `parental level of education`,
## #   2: `test preparation course`, 3: `math score`, 4: `reading score`,
## #   5: `writing score`
```

Seleción de la Muestra para una población

```
set.seed(123456)
df.m<-data.frame(score.marh=sample(df$`math score` ,100,replace = FALSE))
n=nrow(df.m)
head(df.m)
```

```
##   score.marh
## 1          0
## 2         92
## 3         58
## 4         63
## 5         87
## 6         93
```

Se toma una muestra aleatoria de 100; se utiliza la función set.seed() para asegurarnos de que obtengamos los mismos resultados para la aleatorización.

Análisis estadístico de una muestra

Intervalos de confianza

Realizamos un intervalo de confianza del 95% para la variable "math score":

```
test1=z.test(x=df.m$score.marh,
               sigma.x = sd(df.m$score.marh),
               conf.level = 0.95)
test1$conf.int
```

```
## [1] 62.47371 69.12629
## attr(,"conf.level")
## [1] 0.95
```

Prueba Hipótesis

Se indica la prueba hipótesis para la variable "math score":

$$H_0 : \mu = 60$$
$$H_a : \mu \neq 60$$

```
#alternative:"greater", "less" or "two.sided"
#según corresponda la prueba alterna de la hipotesis Ha.
test1=z.test(x=df.m$score.marh,
              sigma.x = sd(df.m$score.marh),
              mu=60,
              alternative="two.sided",
              conf.level = 0.95)
```

##Prueba Hipótesis:resultado

```
test1
```

```
##
##  One-sample z-Test
##
## data:  df.m$score.marh
## z = 3.4176, p-value = 0.0006318
## alternative hypothesis: true mean is not equal to 60
## 95 percent confidence interval:
##  62.47371 69.12629
## sample estimates:
## mean of x
##      65.8
```

Para un nivel de significancia del 5%, podemos afirmar que el valor del Z estadístico es del 3.4175629, además el p-valor es de 6.3184489×10^{-4}.

Prueba Hipótesis: Conclusión

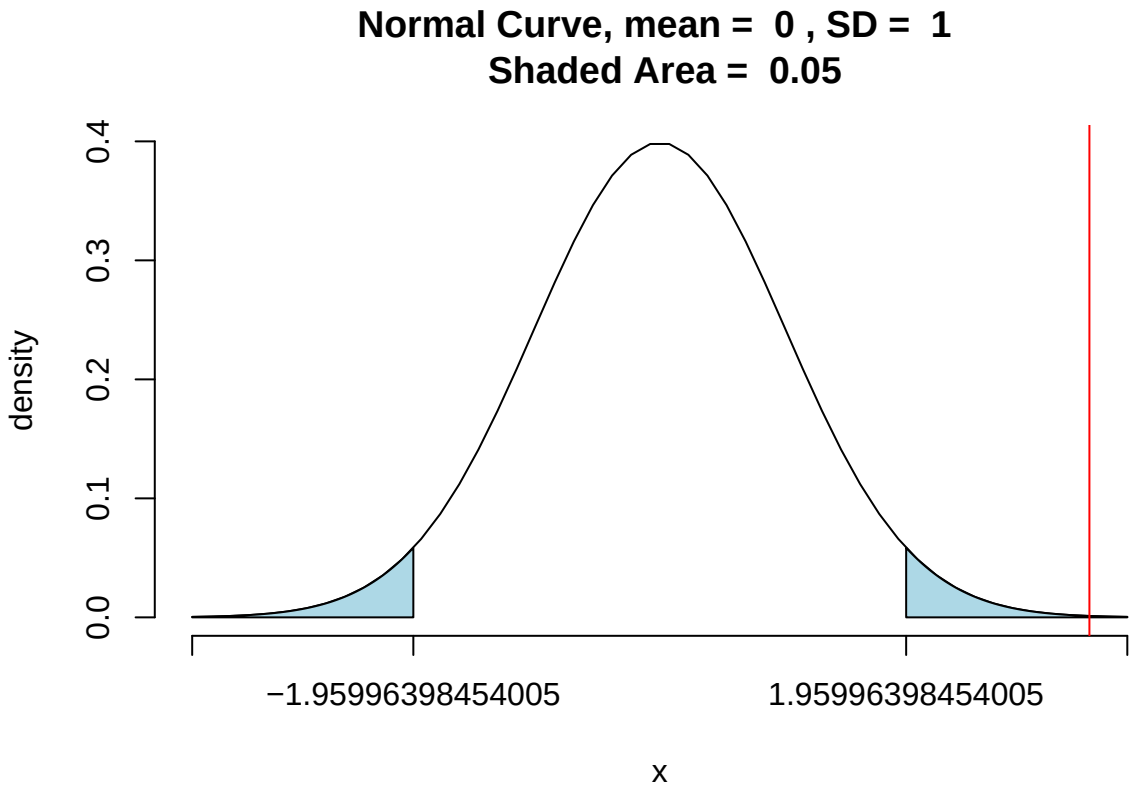
Podemos concluir que existe evidencia suficiente para Rechaza H_0, la media de la variable "math score" no es igual a 65.

```
z1=qnorm(0.025, mean = 0, sd = 1)
z2=qnorm(0.975, mean = 0, sd = 1)
```

```
pnormGC(c(z1,z2),
        region="outside",
        graph=TRUE)
```

```
## [1] 0.05
```

```
abline(v =test1$statistic,
       col="red")
```



Selección de dos Muestra

Se toman dos muestras aleatorias de las variables: “math score” y “reading score”:

```
set.seed(789)
#math score
df.m<-data.frame(score.marh=sample(df$`math score` ,110,replace = FALSE))
#reading score
df.r<-data.frame(score.reading=sample(df$`reading score`,140,replace = FALSE))
```

Intervalos de confianza para la diferencia de dos medias

Realizamos un intervalo de confianza del 95% para la diferencia de medias de las variables “math score” y “score.reading”:

```
test2=z.test(x=df.m$score.marh,
             y=df.r$score.reading,
             sigma.x = sd(df.m$score.marh),
```

```
             sigma.y = sd(df.r$score.reading),
                conf.level = 0.95)
test2$conf.int
```

```
## [1] -7.6004697 -0.4449849
## attr(,"conf.level")
## [1] 0.95
```

El intervalo de confianza nos indica que existe diferencia entre las medias, pues presentan el mismo signo.

Prueba Hipótesis para dos muestras

A continuación se presenta la prueba hipótesis para la direfencias de medias de: $\mu_{score.marh}$ y $\mu_{score.reading}$

$$H_0 : \mu_1 - \mu_2 = 0$$
$$H_0 : \mu_1 - \mu_2 \neq 0$$

```
test2=z.test(x=df.m$score.marh,
             y=df.r$score.reading,
             sigma.x = sd(df.m$score.marh),
             sigma.y = sd(df.r$score.reading),
             mu=0,
             alternative="two.sided",
                conf.level = 0.95)
```

Para un nivel de significancia del 5%, podermos afirmar que el valor del Z estadístico es del: -2.2037362, además el p-valor es de: 0.0275429.

Prueba Hipotesis: Conclusión

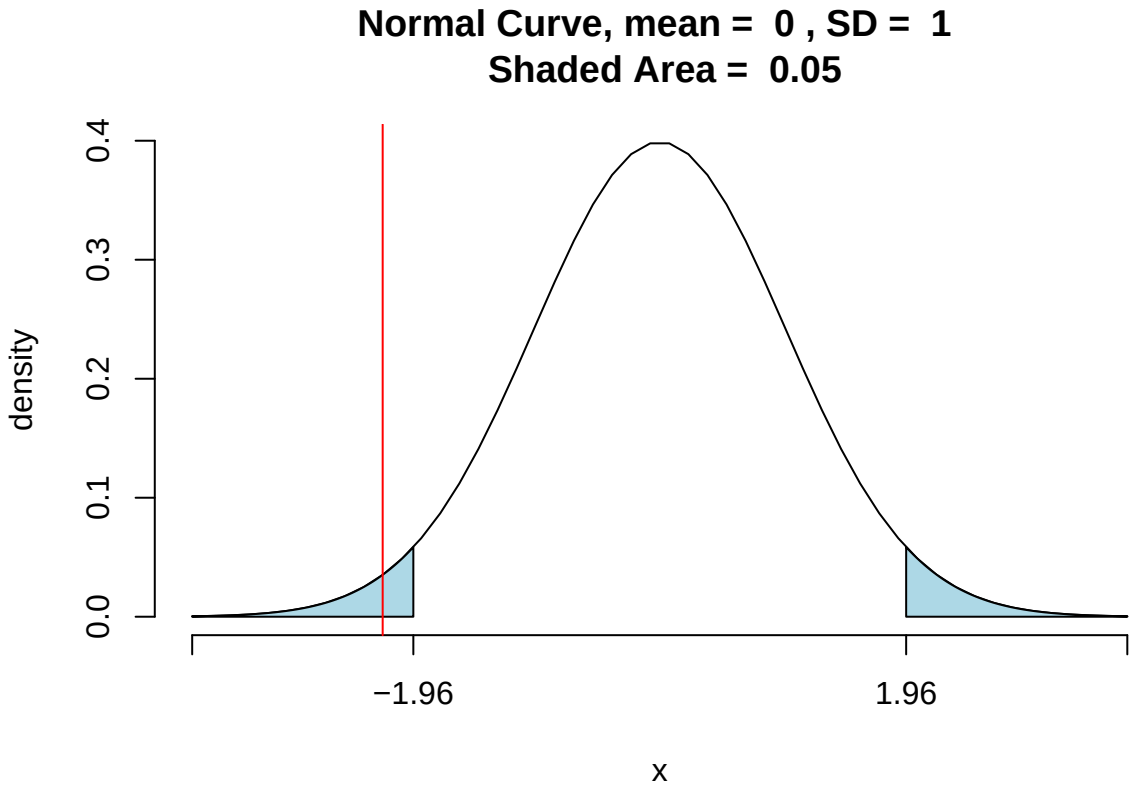
Podemos concluir que existe evidencia suficiente para Rechaza H_0, la diferencia de las medias de las variables "math score" y score.reading no es igual a 0.

```
z1=round(qnorm(0.025, mean = 0, sd = 1),2)
z2=round(qnorm(0.975, mean = 0, sd = 1),2)
```

```
pnormGC(c(z1,z2),
        region="outside",
        graph=TRUE)
```

```
## [1] 0.04999579
```

```
abline(v =test2$statistic,
       col="red")
```



Taller

El taller se entrega de manera individual en la plataforma virtual. El archivo entregable es un HTML donde se debe enetregar código, resultados y ecuaciones aplicadas.

1. Seleccione una base de datos que contenga más de dos variables cuantitativa de escala de medición continua.
2. Seleccione una muestra entre 150 a 200 datos. Cada variable debe tener diferente muestra.
3. De las variables realice un análisis descriptivo.
4. De las variables determine un intervalo del 90% y 95%. De una explicación del resultado.
5. Seleccione una variable y realice una prueba hipótesis de $H_a : \mu \neq \mu_x$. De una explicación del resultado.
6. Seleccione una variable y realice un prueba hipótesis de $H_a : \mu < \mu_x$.De una explicación del resultado.
7. Seleccione una variable y realice un prueba hipótesis de $H_a : \mu > \mu_x$.De una explicación del resultado.
8. Selecciones dos variables continuas y realice un intervalo de confianza para la diferencia de medias. De una explicación del resultado.
9. Realice una prueba hipótesis para la diferencia de medias del 95%. De una explicación del resultado.

bonus extra

- Aplique un intervalo de confianza para una y dos muestras proporcionales. De una explicación del resultado.
- Aplique un intervalo de confianza para una muestras proporcional y para la diferencias de proporciones. De una explicación del resultado.